Network Modeling for High-Dimensional Omics Data

Carel F.W. Peeters[1], Wessel N. van Wieringen[1,2]

[1]Dept. of Epidemiology & Biostatistics, Amsterdam Public Health research institute, VU University medical center Amsterdam, Amsterdam, The Netherlands
[2]Dept. of Mathematics, VU University Amsterdam, Amsterdam, The Netherlands

E-mail for correspondence: `cf.peeters@vumc.nl`, `w.vanwieringen@vumc.nl`

Abstract: Networks are ubiquitous in modern science. Network extraction has become, for many fields, a popular approach to explore a complex of interrelations between entities of interest. The entities in which we take interest are molecular features stemming from omics studies. A challenge with omics-data is that they are are often high-dimensional, i.e., the number of features exceeds the number of observations.

We approach network extraction for high-dimensional data as a problem in penalized graphical modeling. Graphical models utilize graphs to express conditional (in)dependence relations between random variables. Penalization then ensures that these models are estimable from high-dimensional data. We use, in contrast to popular ℓ_1 approaches, an ℓ_2-approach to penalization.

In this course we first show why ℓ_2-based network extraction may be preferred over its ℓ_1-based analogue. We will then focus on the following situations of interest:

- Extracting a single network from steady-state data;

- Simultaneously extracting multiple networks from multiple related data sets and/or data consisting of distinct (disease) subclasses (horizontal integration);

- Extracting networks from time-course data (time-course integration).

Importantly, for each of these situations we will explore methodology to analyze and exploit the networks in order to enhance their practical value. Hence, the course revolves around (i) estimating graphical models, and (ii) translating these models into tangible information and practical consequences for the medical collaborator.

Key words: Graphical modeling; High-dimensional data; Networks; Omics.

Requirements: A laptop, `R`, and the following `R` packages: `rags2ridges` (CRAN, version 2.2), `ragt2ridges` (CRAN, version 0.3.0), `sigaR` (Bioconductor), `lattice` (CRAN), `Biobase` (Bioconductor), `longitudinal` (CRAN), and `SparseTSCGM` (CRAN).

Course materials: `https://github.com/CFWP/MultiOmics-NetworksCourse`, `https://github.com/wvanwie/networkCourse`

Recommended reading:

Bilgrau, A.E., Peeters, C.F.W., et al. (2015). Targeted fused Ridge estimation of inverse covariance matrices from multiple high-dimensional data classes. arXiv:1509.07982v1 [stat.ME].

Miok, V., Wilting, S.M., and van Wieringen, W.N. (2017). Ridge estimation of the VAR(1) model and its time series chain graph from multivariate time-course omics data. Biometrical Journal, 59(1), 172–191.

van Wieringen, W.N., and Peeters, C.F.W. (2016). Ridge estimation of inverse covariance matrices from high-dimensional data. Computational Statistics & Data Analysis, 103(November), 284–303.